Using *i** to Support a Summative Evaluation

James Lockerbie [1], Neil Maiden [1], Amir Dotan [1] & Valentina Lichtner [2]

[1] City University London, Centre for HCI Design, London EC1V 0HB, UK
{J.Lockerbie@soi., N.A.M.Maiden@, Amir.Dotan.1@}city.ac.uk

[2] London School of Economics and Political Science, Information Systems and Innovation Group, London WC2A 2AE, UK
v.lichtner@lse.ac.uk

Abstract. Summative evaluation of a software tool requires the assessment of the defined target outcomes, or high-level goals, of the product. This poses the challenge of how to carry out the assessment in practice. We report our research into addressing this problem by using *i** modelling for a summative evaluation of a work-based learning tool. We describe our use of *i** to identify a set of detailed goals suitable for qualitative assessment. In particular, we report the development and characteristics of the large-scale SR model used in the process, and the utility it provided to contribute towards a successful evaluation. We believe this to be a novel application of *i**, and we present our research outcomes and lessons learned in this area.

1. Introduction

The *i** approach has been widely used in case studies during the early phases of requirements engineering, including our own application to a number of projects as part of our RESCUE process [1]. One such project, called APOSDLE (Advanced Process-Oriented Self-Directed Learning Environment), included *i** modelling to identify future system boundaries, actor dependencies and important system goals for a new knowledge management tool that supports self-directed learning at work. The analysis of this work provided us with a novel insight that the *i** approach would lend itself well to supporting the summative evaluation of the tool at the end of the project.

Summative evaluation of a software tool assesses its defined target outcomes or impacts, and takes place after it has been completely implemented and adequate time has passed to expect outcomes to occur [2]. For APOSDLE, the defined outcomes were expressed as three high-level goals underlying a project vision. Having assessed and evolved the APOSDLE tool itself during two formative evaluations, the aim for the summative evaluation was to assess the satisfaction of these high-level goals to determine whether the product could effectively support learning in the workplace.

2. Objectives of the Research

The main aim of this research was to investigate and evaluate the use of *i** modelling to support a summative evaluation. In particular, our objectives for the study were: (i) to assess the characteristics of an *i** model needed to identify a set of detailed and

measurable goals suitable for qualitative assessment; and (ii) to assess the utility provided by *i** in the evaluation.

In working towards these objectives, we sought to identify lessons learned in order to form an agenda for future work in this area. The outcomes of our research are summarised in the next section.

3. Scientific Contributions

In order to assess the three high-level goals we needed to identify a set of lower level goals suitable for qualitative assessment. Therefore we applied the *i** approach, using models developed in our *i** modelling tool, REDEPEND [3], during four workshops held with project partners. We initially focused on capturing work-based learning soft goals from the application partners, who would later provide the work domains and participants for the summative evaluation. These goals were captured at the start of the project prior to any concrete implementation of the APOSDLE tool, and reflected a more detailed decomposition of the main high level goals – worker support, learner support, and expert support.

We then focused on the input from the technical partners to model potential solution ideas for achieving, or contributing towards, the application partner soft goals. Figure 1 shows the large scale of the SR model, which includes the soft goals of the application partners and the functionality of the APOSDLE system. The expanded section shows an example of the APOSDLE tool (actor B in the figure) contributing towards a non-disturbing learning environment for the knowledge worker (actor A).
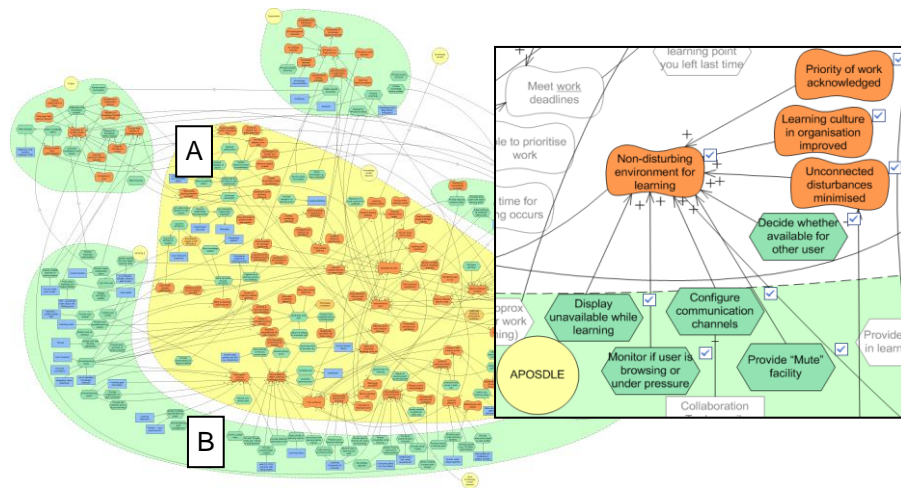

Figure 1: The APOSDLE SR model, with an inset showing functions of the APOSDLE tool and contributions to the knowledge worker soft goals

Based on an assumption of goal hierarchy, lower level goals are more specific than higher level goals and as a consequence lend themselves better to measurement, as

illustrated in Figure 2. Therefore, three goal hierarchies were extracted from the SR model, and the leaf nodes of these hierarchies were taken as goals that could be measured in a meaningful and reliable way. The leaf node soft goals related to the three main aspects of APOSDLE: to support learners, workers and experts in the workplace. As it was not practical to assess all of the low-level soft goals, the application partners identified the ones that were a high priority for the evaluation.
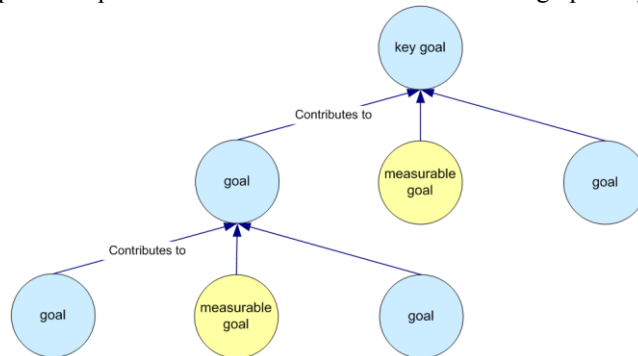

Figure 2: Goal hierarchy showing how lower-level measurable goals can be used to evaluate key high-level goals

Ten key soft goals were selected by the project partners through a questionnaire and follow up meetings as the focus of analysis and evaluation. Qualitative data was collected over a 4 month period, including first-order diary entries [4], interview scripts and log data. Qualitative evidence from the evaluation suggested that the APOSDLE tool contributed towards 9 out the 10 goals, albeit to varying degrees. Given these results, we then explored whether a second-order analysis of the SR model could provide additional insight for the evaluation.

We ran propagations on the prioritised soft goals to identify higher-level soft goals in the hierarchy, and instantiated these parts of the model for each application partner in order to understand the impact of APOSDLE on each of the three domains. It was interesting to find positive contributions applied to a few soft goals that were not supported by the final implementation of the APOSDLE tool. This showed that APOSDLE had system-wide qualities that went beyond the direct implementations intended to achieve application partner soft goals. As expected, the views of the application partners on soft goal achievement varied according to the work domain. It was also interesting to find higher-level goals with positive satisfaction despite lower-level supporting soft goals being reported as unsatisfied.

4.   Conclusions

The research is not complete, but evidence suggests that *i** is an effective tool for structuring a summative evaluation. We assess our two research objectives below.

Our first objective was to assess the characteristics of the *i** model needed to identify a set of measurable goals suitable for qualitative assessment. The lack of clear soft goal hierarchy in the SR model was an issue. We focused on soft goals with the most contributions and flattened out the contributing elements, ensuring that the

majority of the soft goals in the model were covered. A more hierarchical model would have been better suited to the evaluation. Also, it would have been beneficial to have explicitly focused the structure of the SR model on the top level project goals during its development. The interpretation of soft goal descriptions caused problems, with different stakeholders having different understandings, or even no understanding. We later provided rationales for each of the soft goals which improved comprehension. Another challenge for the project partners was the scale of the model, therefore a set of soft goals needed to be prioritised for the evaluation. The project partners had different priorities, and this affected the completeness of the assessment. However, the scale and detail of the model was useful for the analysts, and as such represented a common scalability trade off experienced in *i** modelling.

Our second objective was to assess the utility of applying *i** to the summative evaluation. The main observed benefit was that the SR model provided a set of lower-level goals to structure the evaluation – goals that we otherwise would not have had. These soft goals were also connected to aspects of the tool's functionality, providing context for the evaluation and helping goal selection for the assessment. Highlighting important dependencies and relationships was useful, and showed that the soft goals were not isolated, and that contributing factors propagated throughout the socio-technical system. Whilst the lack of clear hierarchy in the SR model made the identification of measurable soft goals more difficult, this same characteristic of the model also added value to the evaluation. We were able to show that system-wide qualities of APOSDLE went beyond the direct functional implementations intended to achieve application partner soft goals. In addition, we were able to identify contributions that did not fit with the notion of a set goal hierarchy i.e. higher-level goals with positive satisfaction were identified despite lower-level supporting soft goals being evaluated as unsatisfied. Work from this analysis provided additional results and valuable insight for the evaluation.

5.   Ongoing and Future Work

We will take forward these lessons learned for our next project in order to develop more fit-for-purpose models, and to further exploit the observed benefits of applying the *i** approach to summative evaluation.

References

1. Jones S.V. & Maiden N.A.M., 'RESCUE: An Integrated Method for Specifying Requirements for Complex Socio-Technical Systems', In: Requirements Engineering for Socio-Technical Systems, ed. J.L. Mate & A. Silva, Ideas Group, 2005, pp245–265.

2. http://www.evaluationwiki.org/index.php/Summative_evaluation – retrieved on 03/03/2010

3. Lockerbie, J. & Maiden, N.A.M., 'REDEPEND: Tool support for *i** modelling in large-scale industrial projects', Proceedings of the Forum at the CAiSE'08 conference, Montpellier, 2008, pp69–72.

4. Lichtner, V., Kounkou, A., Dotan, A., Kooken, J., Maiden, N.A.M, 'An online forum as a user diary for remote workplace evaluation of a work-integrated learning system', Proceedings of Conference on Human Factors in Computing Systems, 2009, pp2955–2969.